11:00-12:15
1:00-2:15

Week 02 Lecture Notes

How much do we know? The body of human knowledge

June 4, 2019

This week

- Great Paper Chase Due
- "Writing abstracts"

Next week

- Abstract rough draft due
- "The Forgotten Female"
- "What is a Scientific Paper?"

Attendance

- Pass attendance sheet around
- **Be sure to emphasize that it is their responsibility to sign in each week**

Introduction[1]

Last week I had something happen that's never happened to me before in 20 years of teaching. Namely, I missed the first day of class!

I was at a conference for university professors on teaching. It was a fantastic conference and I've come away from it with tons of new ideas, at least some of which I'll likely be trying with y'all this summer.

Fortunately, I have an awesome teaching assistant, Rachel, who was able to get the semester off to a start.

A little about me

Class of '93, graduated with my B.S. in biochemistry in 1994.

Stayed on to get my M.S. in biochemistry in 1999.

Stayed on to teach biochemistry.

Notice any themes here?

Then, in 2008 I had an early midlife crisis and decided to leave A&M to pursue a Ph.D. in evolutionary biology. Write this next bit down because it's important. Ready? Good! When you get to your own midlife crises (a long, long time from now, I know), just go for the sports car and the age inappropriate wardrobe. In the end, it will be FAR cheaper and will take much less of an emotional toll!

[1] Obviously, this material generally goes in my first day notes. This semester, though, I was away at Wakonse for the first lecture so it was necessary to introduce myself.

In any event, I did eventually get my Ph.D. Formally, my degree is in biology, but my dissertation was on cultural, rather than genetic, inheritance, and a big part of cultural inheritance, of course, is writing, which is how I ended up standing before you today.

In 2016, I joined the VIBS faculty to teach VIBS 311 on a part time basis. Indeed, things were so uncertain and so part time that I spent my first year back commuting each week from Dallas to teach rather than move back to College Station full time. I'm extremely happy, though, to have eventually been brought on as a full time faculty member teaching VIBS 311, 310, and some graduate classes in the science and technology journalism program.

Name tents[2]

Give usual spiel about my problems with names.

Have students create name tents. Invite them to include their pronouns if they're comfortable doing so.

Explain that I'll leave them out each week and that they should return them to me.

The body of human knowledge[3]

Last week we introduced the idea of epistemology. Who remembers what we mean by that?

- How do we know what we know?
- How do we distinguish between justified belief and mere opinion?

Today we're going to focus on just how much we know. In other words, we're going to try to wrap our heads around this thing we often call the "body of human knowledge."

So just how big is the body of human knowledge? How much information do we know?

[2] Despite what I tell students, there are actually several reasons for doing this. First, I do, actually, have terrible face-name recognition. Name tents assure that I can call on people by name during class discussions. Students have expressed how meaningful it is to them that I'm at least trying to learn their names, so this has the effect of establishing rapport. Second, whether students share their pronouns or not, asking them to share them signals that my classroom is a safe place for LGBT+ students. Finally, it helps students learn each other's names.

[3] Much of formal science discourse has evolved in response to the need to efficiently access an ever growing body of knowledge (cultural epimemetics, if you will). I've found this point is easier to make with students if I can impress upon them just how big the human body of knowledge is and how fast it's growing.

How would you go about estimating it?

- Get ideas
- One paper I read estimated the number of bytes a human brain holds and then multiplied it by the number of people in the world
- Others used estimates of the capacity of information servers from major corporations such as Google.

This turns out to be an incredibly interesting question with practical importance. It's a question that companies are intensely interested in as it relates to storage, communication, and computation logistics.

And it turns out that there are a lot of different estimates out there, each depending on the assumptions and methods the authors used.

One of the best I've come across is from a paper published in *Science* in 2011 by Martin Hilbert and Priscila Lopez, both from UCLA.[4]

They looked at worldwide estimates based on 60 different categories of information storage spanning two decades.

- 12 families of analog technologies (paper, VHS, records, cassettes, etc.)
- 13 families of digital technologies (hard drives, floppy disks, CD-ROMs, etc.)

Estimated how much of that technology was in use between 1986 and 2007.

They then estimated how much information that represented, normalized for optimal compression.

- This helped correct for some of the redundancy in that information
- However, they did not try to estimate how much of that information was unique

Here are some of their findings:

In 2007, the year they stopped collecting data for the study, they estimated the total amount of information to be 256 EB (exabytes).

That's not a prefix we often run into in our disciplines so let's review.

Kilo	1×10^3
Mega	1×10^6
Giga	1×10^9
Tera	1×10^{12}
Peta	1×10^{15}
Exa	1×10^{18}

So according to their estimates, write the number 256 followed by 18 zeroes!

[4] Hilbert M, López P. The World's Technological Capacity to Store, Communicate, and Compute Information. *Science.* 2011;332(6025):60-65.

Importantly, that number was up from 2.6 exabytes in 1986. That means there was 98.5 times more information in 2007 than there was 21 years earlier. And that number is growing exponentially, so more than a decade later that number today is MUCH bigger!

But even if we know how many zeroes to put behind it, do any of you have a clear idea of just how much information this is?

Here's another way to look at it. In 1986, the total amount of information stored in the world worked out to less than one CD-ROM for every person on the planet. Do any of you remember CD-ROM's? (Take stack of CD-ROMs as a visual.)

By 2007, the amount of information stored worked out to 61 CD-ROMs for every single person on the planet.

Yet another way to look at it is that if you stacked all of those CD-ROMs on top of one another, it would form a tower of CD's from the top of this lectern/desk to the moon… Plus enough to make a quarter of the journey back to Earth!

That's impressive to say the least and, again, it's growing exponentially. This is an accomplishment we as a species can take some pride in. However, it also creates some serious problems that have to be solved.

- No individual can hold all of that information in their head
- That information has to be searchable
- That information has to be accessible

Information is growing so quickly that no individual can keep up with all of it in any detail, not even within very narrow specialties.

One of my areas of scholarship is evolutionary religious studies. How many of you had heard of this discipline before today? It doesn't get much more specialized or obscure than that. Yet as narrow as it is, I've never been able to keep up with it all.

<u>Information acquisition and retrieval</u>

So let's turn our attention now to some of the problems associated with having so much information. Last week I assigned "The Great Paper Chase." How was that? Was it easy? Challenging?

Very few of you contacted me so I'm assuming it went reasonably well. What problems did you run into as you worked on it?

- Get feedback from students.

What methods did you use to find your answers?

- Again, get feedback.

This is going to be an extremely important skill for you this semester and beyond. I strongly suggest that if you have any questions about how to do this that you seek help either from the libraries on campus or from me and Rachel.

Abstracts[5]

Abstracts are one of the many tools scientists use to address these problems of information overload that we've been talking about today.

Let's start with the term itself, though. What usually comes to mind when someone uses the word "abstract?"

- Probably something that exists in theory or in the mind but that doesn't have physical, material existence.

However, "abstract" also has another meaning, one that isn't used much in everyday language.

- Namely, "abstract" also means to extract something, distilling it into its most essential parts.

An abstract, then, is a summary of a paper, a summary that provides the most important information contained in a paper.

We'll be talking about the structure of scientific papers next week and you'll see that most abstracts have the same structure. The difference is that an abstract compresses all of that information into a single paragraph.

As with anything when we're talking about writing, there are exceptions to this structure.

- Some journals, for instance, label each section separately.
- Some rearrange the order
- Some include other kinds of information

The vast majority, however, will adhere to the following structure:

Title

We aren't going to worry about titles for this assignment because, for journal articles, the title of the abstract is the same as the title of the article.

[5] This part of the lecture supports the reading I had students do. I return repeatedly to the structure of scientific articles throughout the semester. Abstracts help students learn and internalize that structure. Note that, at some point, I will probably rewrite my primer on writing abstracts so that my labels for each section match Gastel's IMRAD labels.

Even so, as you were finding articles for The Great Paper Chase, you probably read more than a few titles. How would you characterize them?

- Long
- Dry
- Boring
- Descriptive
- Very detailed and specific
- Usually in two parts separated by a colon

None of them are generally as catchy as *A Tale of Two Cities* or *Game of Thrones*. Why?

- Because catchy titles don't usually reveal content
- Remember that one thing titles need to do is address the problem of information retrieval
- Detailed titles help you sort through all the excess information to get to what you need

Background

Continuing with that theme of information retrieval, the first sentence or two of an abstract will provide the context for the study, the background information needed to understand how this study fits in with the rest of the body of human knowledge.

Imagine that stack of CD-ROMs stretching from here to the moon and back.

You need some kind of "you are here" sign to know just where in that stack of CD-ROMs you're looking.

If you're looking for an article about vector borne diseases in the tropics and you read in the first sentence that the paper is looking at vectors used to transfer DNA between bacteria, then you know you don't need to read anything more. You've got the wrong paper.

Just consider the elegance of that. The detailed title gives you an indication of whether an article is potentially useful. Just two sentences into the abstract, you probably already make a decision about whether it's worth your time to read more.

How much time do you think that saves?

Objective

Just after the background information, most abstracts state what the specific objectives of the study were. What gap are they trying to fill? Just what problem were they trying to address? What specific questions were they asking? In short, why did they undertake this research project and what did they hope to find?

Methods Used

You've all been to restaurants and read menus. What information do menus provide?

- What dishes are on offer
- How much they cost
- What each dish contains

Do menus provide recipes for preparing the dish yourself? Of course not!

The methods section of an abstract is very much the same way. It lists the methods used, the experimental approach taken, but it does not list enough detail that readers could repeat the experiments from the abstract alone.

Results

Next up after the methods are the major results of the study. It's important to remember that most studies generate a lot of data. The abstract does not report on every single bit of information! Instead, it only highlights the most important results, the findings that the authors used to justify their conclusions.

An important thing to keep in mind is that results are just facts. They can't be argued. They can't be debated. Your team either won, lost, or tied the game last week. The score was the score. Facts are, however, open to interpretation, which is what the last section of the abstract is for.

Conclusions

The conclusions section of the abstract is where the authors offer their interpretations of the results. Your team lost last week? Why did it lose? Was it because the offense was weak? Or was it a fault of the defense? These are conclusions drawn from the facts of the game. Some conclusions are stronger than others, better supported by those facts. Nevertheless, they are interpretations and other people might look at those same facts and interpret them differently, coming to different conclusions.

Just keep in mind that results are facts. Conclusions are interpretations. And authors usually offer their interpretations after they present the facts.

So that, in a nutshell, is an abstract. Each section is usually one or two sentences long. Maybe three. Again, though, note the elegance of that paragraph. It can take a long time to read an entire paper. It doesn't, however, take long to read a paragraph.

Not only does this help you find the information you need for your work, it is also an excellent way of coming up to speed on a new topic. Yes, eventually you should read full papers. However, in the beginning you can learn an awful lot very quickly by just reading abstracts.

Abstract Practice[6]

I want to give you some practice applying what you learned from the reading with some practical examples.

Pass out abstracts. [I selected 5 abstracts from different disciplines from my Endnote library and printed them out, collating them so that they would be more or less evenly distributed around the room to form 5 groups.]

The first thing I'd like you to do is read your abstract and I'd like you to mark what you think are the:

- Title
- Background
- Objective
- Methods
- Results
- Conclusions

Give them about 5 minutes to complete this task. Watch the time!

Okay, I passed out 5 different abstracts. I want you to find the other students with the same abstract as you.

- If time permits, have them discuss with each other what they thought each study was about.
    - What gap did this study fill?
    - What was not known before the study started?
    - What was known?
- If time is short, have students compare their labels.

So, did y'all identify the same things?

If not, what were some of the differences?

Did any group's abstract not have all of these parts?

Open to general discussion, drawing out the following points.

- The abstract is used in nearly every scholarly discipline
- Not all abstracts have exactly the same structure
- However, nearly all of them will convey the same information

[6] I make extensive use of this active learning technique in my writing courses. I don't know if it has a formal name, but it's essentially a variation of the think/pair/share method. My inspiration for this technique came from the structure of Wakonse South conferences, in which participants are assigned dialogue groups they return to throughout the conference in order to discuss focused topics and to share what they learned from concurrent sessions. I've found that it generates a lot of class discussion, even in classes where students are reluctant to speak up. It also helps establish rapport in the class because students get to talk with students who aren't their immediate neighbors.

11:00-12:15
1:00-2:15

Assignment 1[7]

Pass out writing prompt

Go over it in as much detail as time will allow.

An especially important point to go over is that their TA will be providing them specific goals to work on for the final draft. They must write those goals at the top of their abstract so I can assess whether they have made progress on those goals.[8]

Reading for next week:[9]

- "The Forgotten Female"
- "What is a Scientific Paper?"

[7] See http://hdl.handle.net/1969.1/175371. In this assignment, I provide three old research articles that were published without abstracts. Each student choses one of those articles and writes an abstract for it. Students turn in a rough draft of the article, which is then reviewed by their teaching assistant and three of their peers. They then revise their rough draft based on the feedback they receive before turning in the final draft for me to grade.

[8] This is something new I'm trying this semester. I attended a concurrent session at Wakonse about providing feedback on student writing. How I've been approaching it up to now is unsustainable. It takes too long for me to get feedback to students, sucks up almost all of my waking hours, and, in the end, I'm not convinced it's really giving them the information they need to improve. This approach is a slight variation of an approach advocated for at Wakonse. I'm trying it because I think it addresses two issues. First, I'm hopeful it will allow me to provide more useful feedback with less effort. Second, it will force students to review their TA's comments, something many of them neglect to do, especially for the first assignment.

[9] These readings set up my next lecture, in which I attempt to make the point that scientific discourse is just an elaborate, formalized form of storytelling.

One

Latitudinal patterns and environmental determinants of recent human cultural diversity: do humans follow biogeographical rules?

Biogeographers have noted many strong patterns in the diversity and distribution of animal and plant taxa. Human cultural diversity also exhibits strong geographical patterns. Here we analyse the global distribution of 3814 human cultures in relation to latitude and climatic parameters. The density and diversity of human cultures decline with latitude and increase with temperature and rainfall. Human cultures in tropical, wetter or warmer areas have smaller ranges and are more densely packed and differentiated. These relationships can be documented statistically in ways that parallel species diversity among other organisms. The global nature of these patterns implies ecological equilibrium independent of evolutionary history in different continents. This has implications for the interpretation of human genetic diversity, as well as for the understanding of processes of human cultural diversification and their relationship to evolutionary and ecological mechanisms.

Two

Nobody's watching? Subtle cues affect generosity in an anonymous economic game

Models indicate that opportunities for reputation formation can play an important role in sustaining cooperation and prosocial behavior. Results from experimental economic games support this conclusion, as manipulating reputational opportunities affects prosocial behavior. Noting that some prosocial behavior remains even in anonymous noniterated games, some investigators argue that humans possess a propensity for prosociality independent of reputation management. However, decision-making processes often employ both explicit propositional knowledge and intuitive or affective judgments elicited by tacit cues. Manipulating game parameters alters explicit information employed in overt strategizing but leaves intact cues that may affect intuitive judgments relevant to reputation formation. To explore how subtle cues of observability impact prosocial behavior, we conducted five dictator games, manipulating both auditory cues of the presence of others (via the use of sound-deadening earmuffs) and visual cues (via the presentation of stylized eyespots). Although earmuffs appeared to reduce generosity, this effect was not significant. However, as predicted, eyespots substantially increased generosity, despite no differences in actual anonymity; when using a computer displaying eyespots, almost twice as many participants gave money to their partners compared with the controls. Investigations of prosocial behavior must consider both overt information about game parameters and subtle cues influencing intuitive judgments.

Three

Protein stability promotes evolvability

The biophysical properties that enable proteins to so readily evolve to perform diverse biochemical tasks are largely unknown. Here, we show that a protein's capacity to evolve is enhanced by the mutational robustness conferred by extra stability. We use simulations with model lattice proteins to demonstrate how extra stability increases evolvability by allowing a protein to accept a wider range of beneficial mutations while still folding to its native structure. We confirm this view experimentally by mutating marginally stable and thermostable variants of cytochrome P450 BM3. Mutants of the stabilized parent were more likely to exhibit new or improved functions. Only the stabilized P450 parent could tolerate the highly destabilizing mutations needed to confer novel activities such as hydroxylating the antiinflammatory drug naproxen. Our work establishes a crucial link between protein stability and evolution. We show that we can exploit this link to discover protein functions, and we suggest how natural evolution might do the same.

Four

Sacred groves: potential for biodiversity management

Existing global protected area networks have two shortcomings: (1) they do not cover certain habitats, and (2) local people often resent their formal management. Here, we show that communities around the world traditionally protect natural sites that are dedicated to ancestral spirits or deities. Such sites cover a wide variety of habitats and are often located in biodiversity rich regions. Case studies on sacred groves show that these small forest patches play an important role in biodiversity conservation. Furthermore, natural sacred sites are maintained through traditional methods of community based conservation that do not require governmental involvement. Incorporating these sites into conservation networks could enhance the effectiveness of protected areas by covering a wider variety of habitats and by harnessing the support of local people. In this article, we discuss current threats to sacred groves that need to be addressed through management approaches. More research on the ecology and underlying socioeconomic mechanisms of natural sacred sites is required to fully reveal their potential for biodiversity conservation.

Five

Zika virus in the Americas: Early epidemiological and genetic findings

Brazil has experienced an unprecedented epidemic of Zika virus (ZIKV), with ~30,000 cases reported to date. ZIKV was first detected in Brazil in May 2015, and cases of microcephaly potentially associated with ZIKV infection were identified in November 2015.We performed next-generation sequencing to generate seven Brazilian ZIKV genomes sampled from four self-limited cases, one blood donor, one fatal adult case, and one newborn with microcephaly and congenital malformations. Results of phylogenetic and molecular clock analyses show a single introduction of ZIKV into the Americas, which we estimated to have occurred between May and December 2013, more than 12 months before the detection of ZIKV in Brazil. The estimated date of origin coincides with an increase in air passengers to Brazil from ZIKV-endemic areas, as well as with reported outbreaks in the Pacific Islands. ZIKV genomes from Brazil are phylogenetically interspersed with those from other South American and Caribbean countries. Mapping mutations onto existing structural models revealed the context of viral amino acid changes present in the outbreak lineage; however, no shared amino acid changes were found among the three currently available virus genomes from microcephaly cases. Municipality-level incidence data indicate that reports of suspected microcephaly in Brazil best correlate with ZIKV incidence around week 17 of pregnancy, although this correlation does not demonstrate causation. Our genetic description and analysis of ZIKV isolates in Brazil provide a baseline for future studies of the evolution and molecular epidemiology of this emerging virus in the Americas.